SYNERGY PUBLISHING PTE. LTD.

Modern Electronic Technology

https://ojs.s-p.sg/index.php/met

An Improved NID Steganography Method for Low Bitrate Speech on VoIP

Jin Liu* **Yiwen Zhang**

College of Computer Science and Technology, National Huaqiao University, Xiamen, Fujian, 361021, China

ARTICLE INFO

Article history
Received: 14 January 2021
Revised: 21 February 2021
Accepted: 9 April 2021
Published Online: 16 April 2021

Keywords:
Low Bitrate Speech
Steganography
Voice Over IP
QIM

ABSTRACT

In the procedure of encoding process on low bitrate speech, fixed codebook division is an efficient and promising embedding method for steganography. An improved neighbor index division (NID) steganography method based on the high bitrate frame of G.723.1 codec (6.3kbit/s) is proposed, which employs the parity and low distortion of neighbor indices for G.723.1 fixed codebooks. Differing from previously NID method which performs quantized index modulation (QIM) beforehand, the proposed method divides codeword indices into separate sub-codebooks according to the secret message bits dynamically in the original G.723.1 codec quantization period. Compared with existing NID method, our proposed method doesn't need to divide the codebook before the encoding starts. The embedding and codebook dividing happen simultaneously, which utilizes the characteristics of specific secret message bits. The experiment results show that the proposed method has a much lower quality degradation for the decoding speech and still fulfills the low latency requirement for communication.

1. Introduction

With the rapid growth of traffics on the Internet, especially on cloud computing and Internet of Things, the opportunity of steganography on various covers has come, such as the speech flow for Internet telephone services. Voice over IP (VoIP) is one type of main service among them, which becomes an appropriate type of covert carrier for secret communication [1]. Compared with other types of steganography covers, the low bit-rate speech in VoIP has many advantages. The first is the huge amount of flow it brings for communication, which always means more embedding capacities. And on the other hand is the instantaneity of it, which left limited analyzing time for a malicious listener[2]. From another point of view, the low bitrate feature makes the redundancy in speech covers limited, which is the fundamental for steganography.

Many steganography methods have been published for low bitrate speech, which can be roughly divided into two groups. The first one takes advantage of the networking protocols underlying VoIP service, such as RTP/RTCP[3]. However, the restricted segment of the protocol headers makes this type of steganography vulnerable, and always has a low embedding capacity[4]. The other type exploits the huge amount of redundancies in the abundant speech payloads of VoIP packets. The commonly used speech codecs are SILK, ITU-T G.729, ITU-T G.723.1[5], IETF iLBC etc. Within this type of steganography, the embedding methods includes variant least significant bit (LSB)[6], speech bitrate switch[7] and quantized index modulation (QIM)[8]. This type of methods always bring less change

**Corresponding Author:*
Jin Liu,
College of Computer Science and Technology, National Huaqiao University, Xiamen, Fujian, China;
E-mail: geneleo@hqu.edu.cn.

Distributed under creative commons license 4.0 DOI: https://doi.org/10.26549/met.v5i1.6167

to the flow feature, yet influence the receiving speech quality.

The QIM method in the second group employs the internal feature of speech codec, which divides the quantization codebook into separate groups (sub-codebooks). In the speech encoding procedure the groups in which the encoding algorithm searches for parameter quantization depends on the secret bits for transmission. For this reason the performance of the steganography methods largely lie on the codebook division approaches, which may influence the speech distortion after decoding and the embedding capacity of covert communication. Previously proposed methods[8] often use offline algorithms to divide the codebooks beforehand, which low the amount of processing time for steganography. However, the offline algorithms always base on the average tested speech loss and introduced secret bits, the real time speech and current secret bits may differ from it. As a result, taking advantage of the individually introduced characteristics to each frame could reduce the distortion for the speeches. The rest of the paper is organized as follows.

In section 2, we briefly discuss the quantization procedure of G.723.1 speech codec and the existing neighbor index modulation (NID) steganography. The proposed improved NID steganography is given in section 3. Then section 4 gives the experiment and performance analysis of it. In the end, section 5 concludes the whole paper.

2. G.723.1 Speech and NID Steganography

G.723.1 is a speech encoding standard specifies in ITU-T Recommendation[9], which compresses the original PCM speech into 6.3kbit/s frames(low bitrate) and 6.3kbit/s frames(high bitrate). G.723.1 belongs to parameter based speech coding method, which encodes PCM speech into frames made up of fixed frame parameters. It is suitable for real time VoIP applications especially at bandwidth limited circumstances. QIM steganography mainly focuses on the quantization stage of line spectral pair (LSP) coefficients which consists of three separate codebooks. Table 1 shows the three fixed codebooks for LSP quantization, where the dimension value denote the number of values for each codeword (vector). Codebook$_i$ in Table 1. is made up of 256 codewords with an index ranging from 0 to 255.

Table 1. Fixed Codebooks of G.723.1 6.3kbit/s Frame

LSP Quantization	Index Length(bits)	Codewords	Dimension
Codebook$_0$	8	256	3
Codebook$_1$	8	256	3
Codebook$_2$	8	256	4

QIM based methods divide each codebook into separated sub-codebooks, which construct different redundant states for input speech on the LSP quantization procedure, thus secret bits could be embedded accordingly into the indices of sub-codebooks. Figure 1 gives a brief description of the embedding procedure based on LSP quantization.

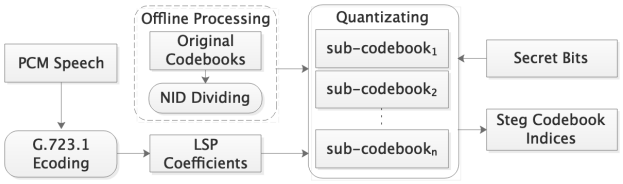

Figure 1. QIM and NID algorithm for G.723.1 codec

The key point for Fig.1 is the codebook division algorithm, which includes CNV[10], NID[8], etc. In NID steganography, the codebooks are divided according to the parity of the codeword indices. The codewords in each codebook are separated into n sub-codebooks offline before the encoding starts, where n depends on the specific embedding capacity of each codebook needed for each cover communication session. That is, the secret bits are embedded based on the value of indices into which the LSP coefficients are quantized.

3. Improved NID Steganography Method

For both NID and CNV codebook dividing algorithms, the main principle is to separate the nearest indices into different sub-codebooks. However, the search process for all nearest neighbor indices in CNV[10] has a time complexity about O(n^n), where n is the number of indices. It is a time-consuming process, so the codebook dividing for CNV method is always performed aforehand before the G.723.1 encoding starts. When it comes to NID method, the indices are separated inside or outside of the G.723.1 codec, because the codebook division operation has no relationships with the vector quantization process. NID is only responsible for separating adjacent indices into different sub-codebooks. For instance, if index$_i$ is supposed to be classified into sub-codebook$_i$, then its adjacent index$_{i-1}$ and index$_{i+1}$ must be put into the sub-codebook(s) differing from sub-codebook$_i$.

As the NID method doesn't take the quantization into account, the onsite secret bits will affect the LSP quantization process, thus influence the restored speech quality on the receiver side. Instead, our proposed method uses original searching process, and searches more indices (l >=1) for steganography, which has at least two advantages. On one hand, the search is performed online, on the worst condition the second optimal index is used for steg-

Distributed under creative commons license 4.0
DOI: https://doi.org/10.26549/met.v5i1.6167

anography. So that the distortion will be reduced to the limit. On another hand, the additional information, such as the previously divided sub-codebooks, will not needed to be transferred to the receiver, which results flexible covert communication. For the proposed improved NID algorithm, we set l to 2. It means that the resulted number of indices for the search algorithm is two (v_i and $v_{i'}$), where v_i is the optimal index for the search algorithm of G.723.1 and $v_{i'}$ is the second optimal one. The main step for the search operation for LSB quantization is described as follows.

For the input LSP coefficients p_i and secret bit b_i, if $b_i == 0$ and $v_i \bmod 2 == 0$ then return v_i, else if $b_i == 1$ and $v_i \bmod 2 == 1$, return v_i, else if $b_i == 1$ and $v_i \bmod 2 == 0$ and $v_{i'} \bmod 2 == 1$, return $v_{i'}$, else if $b_i == 1$ and $v_i \bmod 2 == 0$ and $v_{i'} \bmod 2 == 0$, return $v_i - 1$ or $v_i + 1$ (use alternative neighbor index). The embedding process is illustrated in Figure 2, where all the three fixed codebooks of G.723.1 speech codec are used and the steganography bits are shown as examples of different cases. The key feature is that the steganography operation is performed within the original G.723.1 codec, additionally, one more optimal index is searched out compared with the primitive quantization.

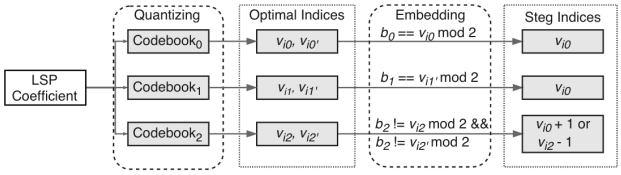

Figure 2. Improved NID Steganography

In the improved NID method, optimal and second optimal indices are obtained by real time search operation, which utilizes the characteristics of each LSP parameter to be quantized. In this regard, the steganography operation is optimized compared with the previously divided sub-codebook methods, which only take the average parameter features into account. The introduced time consumption is low as only 2 optimal indices are searched and the processing time for this operation takes up a low percentage in the G.723.1 encoding process. The time complexity for searching is O(n), and the second optimal index is get by simply excluding the optimal index when performing a second search operation or searching for two optimal indices in a single searching function. The first way is much easier on implementation, and the second way is more time efficiency. If the two optimal indices doesn't fulfill the embedding requirements, then NID method is used. That is, use the adjacent index as the result (steg codeword index).

4. Experiment and performance analysis

Because of the optimal searching for vertices of the proposed method, the speech quality after embedding outperforms previous QIM based method, such as NID and CNV. To demonstrate it, we adopt perceptual evaluation speech quality (PESQ)[11] criterion to evaluate the degraded speech quality after steganography. PESQ is an ITU-T Recommendation for objectively evaluating speech quality, which can be translated into Mean Opinion Score—Listening Quality Objective (MOS-LQO) ranged from 1.017 to 4.549. The sample datasets are collected from news reports and daily conversations, which include four categories: Male Mandarin (MM), Female Mandarin (FM), Male English (ME) and Female English (FE).

Table 2. Speech quality evaluation using PESQ

Samples	MOS-LQO			
	G.723.1	CNV	NID	Improved NID
MM	3.522	3.488	3.481	3.501
FM	3.405	3.329	3.326	3.358
ME	3.324	3.302	3.296	3.311
FE	3.265	3.237	3.233	3.242
Mean Value	3.379	3.339	3.334	3.353
Decrease (%)	-	1.18	1.33	0.77

Table 2 gives the PESQ evaluation results of G.723.1, CNV, NID and our proposed improved NID. The G.723.1 column evaluates the speech degrading after the original G.723.1 encoding and decoding, where the input are the sample PCM speech and decoded speech after the decoding of received G.723.1 frames. We give the translated MOS-LQO value of them, which is more intuitive. The results show that the improved NID steganography has a better speech quality, which is only decreased with 0.77% compared with original G.723.1 speech. In addition, to fulfill the real time requirement for communication, we also tested the latency introduced by the proposed method. The additional latency introduced for MM, FM, ME and FE are less than 35 in average on an Intel Pentium E5200 processor with the 2.5 GHz frequency, which is much less than the G.723.1 encoding duration about 800. Moreover, compared with the 150ms end to end latency requirement for communication, the introduced latency by improved NID steganography is negligible.

5. Conclusion

An improved NID steganography method is proposed in this paper, which utilizes the inherent search process of G.723.1 quantization. The embedding operation is per-

Distributed under creative commons license 4.0 DOI: https://doi.org/10.26549/met.v5i1.6167

formed within the codec, so that it is almost the lowest distortion introduced for steganography among those QIM based methods. The experimental results show that this method outperforms previous CNV and NID methods, yet still fulfills the real time requirement of communication.

6. Acknowledgements

This work is supported in part by the First Batch of Youth Innovation Fund Projects in 2020 under Grant No.3502Z202006012 and the Experimental Teaching Reform Project of National Huaqiao University under Grant No.SY2019L013.

References

[1] Peng J, Tang S. Covert Communication over VoIP Streaming Media with Dynamic Key Distribution and Authentication[J]. IEEE Transactions on Industrial Electronics, 2020.

[2] Yang H, Yang Z, Huang Y. Steganalysis of voip streams with cnn-lstm network[C]//Proceedings of the ACM Workshop on Information Hiding and Multimedia Security. 2019: 204-209.

[3] Azadmanesh M, Mahdavi M, Ghahfarokhi B S. A reliable and efficient micro-protocol for data transmission over an RTP-based covert channel[J]. Multimedia Systems, 2019: 1-18.

[4] Zhang X, Tan Y A, Liang C, et al. A covert channel over volte via adjusting silence periods[J]. IEEE Access, 2018, 6: 9292-9302.

[5] Kabal P. ITU-T G. 723.1 speech coder: A matlab implementation[J]. McGill Univ, 2004.

[6] Yang W, Tang S, Li M, et al. Markov bidirectional transfer matrix for detecting LSB speech steganography with low embedding rates[J]. Multimedia Tools and Applications, 2018, 77(14): 17937-17952.

[7] Liu J, Tian H, Zhou K. Frame-bitrate-change based steganography for voice-over-IP[J]. Journal of Central South University, 2014, 21(12): 4544-4552.

[8] Liu J, Tian H, Lu J, et al. Neighbor-index-division steganography based on QIM method for G. 723.1 speech streams[J]. Journal of Ambient Intelligence and Humanized Computing, 2016, 7(1): 139-147.

[9] ITU-T Rec. G.723.1. Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s[S], 2006.

[10] Tian H, Liu J, Li S. Improving security of quantization-index-modulation steganography in low bit-rate speech streams[J]. Multimedia systems, 2014, 20(2): 143-154.

[11] ITU-T Rec. P.862. Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs[S]. 2001.

Distributed under creative commons license 4.0
DOI: https://doi.org/10.26549/met.v5i1.6167